Contrast Invariant SNR

Weiss, Pierre; Escande, Paul ; Dong, Yiqiu

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Contrast Invariant SNR

Pierre Weiss, Paul Escande and Yiqiu Dong

***Abstract*—We design an image quality measure independent of local contrast changes, which constitute simple models of illumination changes. Given two images, the algorithm provides the image closest to the first one with the component tree of the second. This problem can be cast as a specific convex program called isotonic regression. We provide a few analytic properties of the solutions to this problem. We also design a tailored first order optimization procedure together with a full complexity analysis. The proposed method turns out to be practically more efficient and reliable than the best existing algorithms based on interior point methods. The algorithm has potential applications in change detection, color image processing or image fusion. A Matlab implementation is available at http://www.math.univ-toulouse.fr/~weiss/PageCodes.html.**

***Index Terms*—Local contrast change, topographic map, isotonic regression, convex optimization, illumination invariance, signal-to-noise-ratio, image quality measure.**

I. Introduction

INVARIANCE to illumination conditions is often a key element for the success of image processing algorithms. The whole field of mathematical morphology is based on contrast invariance [1]. The structural similarity index [2] - one of the most popular image quality measures - also strongly relies on a partial invariance to illumination changes.

In this paper, we introduce a novel algorithm that allows comparing two images in a way independent of *local contrast changes*. Most of the works dedicated to illumination invariance consist of decomposing the image into patches and to normalize means and variances on the patches. Here, we follow a different trail and consider that two images differ by a local contrast change if they share the same *component tree* [3], [4]. Given a reference image $u_0 : \Omega \to \mathbb{R}$ and another image $u_1 : \Omega \to \mathbb{R}$, we propose to measure their similarity by computing the following value:

$$\Delta_{loc}(u_1, u_0) = \min_{T \in \mathcal{T}} \|u_0 - T(u_1)\|_2^2, \tag{1}$$

where $\mathcal{T}$ is the set of local contrast changes. This amounts to finding the best match with u_0 among all images that have the same component tree as u_1. The locally contrast invariant signal-to-noise-ratio is defined by:

$$SNR_{loc}(u_1, u_0) = -10 \log_{10}(\Delta_{loc}(u_1, u_0)/\|u_0\|_2^2). \tag{2}$$

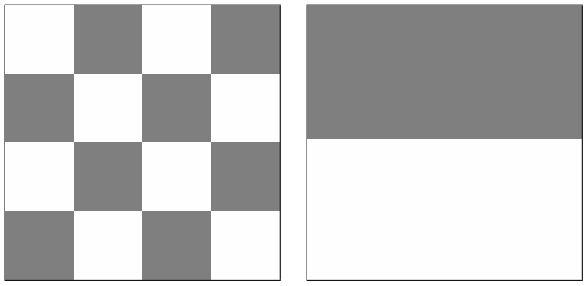
Let T^* denote the optimal contrast change in problem (1). The image $u^* = T^*(u_1)$ has a geometry identical to u_1, with the contrast of u_0.

We introduce an efficient algorithm based on convex programming to solve (1) and provide a full analysis of its complexity. The proposed method has potential applications in different fields such as image quality measure, change detection [5], image fusion or color image processing [6]–[8].

Y. Dong is with the department of applied mathematics and computer science at Technical University of Denmark.

P. Escande is with the DISC department, at ISAE, Université de Toulouse, France.

P. Weiss is with ITAV and IMT, CNRS and Université de Toulouse, France.



(a) (b)

Fig. 1: Two images with different geometrical features can have an identical histogram.

II. Existing approaches

Various approaches are commonly used to compare two images u_0 and u_1 independently of illumination variations. We briefly describe a few of them below. A Matlab implementation of each method is provided here http://www.math.univ-toulouse.fr/~weiss/PageCodes.html.

A. Contrast equalization

Probably the most common approach consists of equalizing histograms, i.e. to change the gray-values of u_1 in such a way that the resulting histogram matches approximately that of u (see e.g. [9], [10]). This approach suffers from the fact that the image geometry is completely forgotten: histograms only account for gray-level distributions and not geometrical features such as edges, textures,... An example of two images with identical histogram and different geometrical contents is provided in Fig. 1.

B. Optimal linear and affine maps

The set $\mathcal{T}$ in problem (1) can be replaced by any class of transformations that describe changes of illuminations. Probably the simplest classes $\mathcal{T}$ are the set of linear maps $T(u) = au$ or the set of affine maps $T(u) = au + b$, where a and b are scalars. The solution of both problems can be computed explicitly in terms of u_0 and u_1. The same approach can be used locally and the L^2-norm can be replaced by a weighted L^2-norm. This idea is the basis of the Structural Similarity Index Measure (SSIM).

C. Optimal global contrast change

A richer set of transformations $\mathcal{T}$ is that of global contrast changes. Two images u_0 and u_1 are said to differ by a global contrast change if there exists a non increasing function $T : \mathbb{R} \to \mathbb{R}$ such that $T(u_1) = u_0$. Finding the best global contrast change amounts to solving:

$$\Delta_{glo}(u_1, u_0) = \min_{T:\mathbb{R}\to\mathbb{R},\ \text{non increasing}} \frac{1}{2}\|T(u_1) - u_0\|_2^2. \quad (3)$$

We let

$$SNR_{glo}(u_1, u_0) = -10\log_{10}(\Delta_{glo}(u_1, u_0)/\|u_0\|_2^2)$$

denote the globally contrast invariant SNR. Problem (3) is a simple case of isotonic regression [11], [12]. Due to the simple structure of the constraint set, this problem can be solved in $O(n)$ operations using active sets type methods.

Unfortunately, global contrast changes do not capture all the complexity of illumination changes: in most applications, the variations are local.

III. Optimal local contrast changes

A. Topographic maps and local contrast changes

Mathematical morphology emerged with the works of Matheron [13]. Therein, he proposed to analyze an image u by using geometric operations on its (upper)-level sets $\chi_\lambda = \{x \in \Omega, u(x) \geq \lambda\}$ or its level lines $LL_\lambda = \{x \in \Omega, u(x) = \lambda\}$[1]. The level-sets and level lines are geometrical features invariant to global contrast changes. In order to obtain a representation invariant to local contrast changes, it is possible to consider the *connected components* of these objects. This idea was proposed and detailed thoroughly in [3], [14]. The connected components, thanks to the inclusion relationships, can be encoded in a tree structure called component tree, useful in many applications [4], [15]. This tree structure allows formalizing the notion of local contrast change: two images differ by a local contrast change if they share the same component tree.

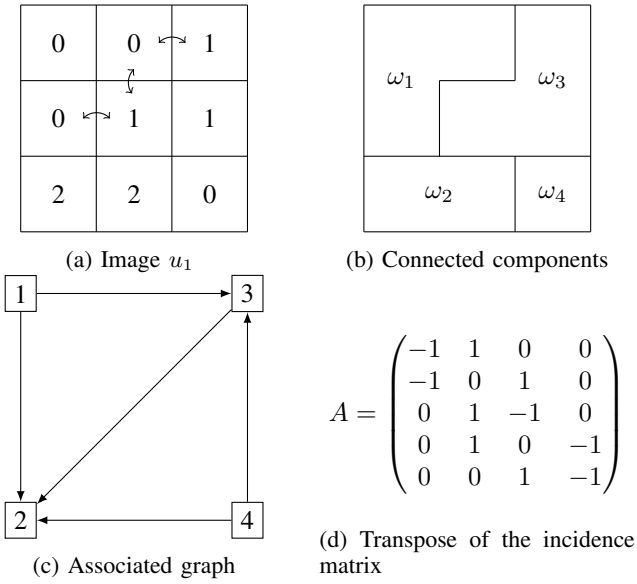
This definition has many assets. In particular, it allows working both in the discrete and continuous settings. For our purposes however, we will use a simpler definition valid for discrete images only. Let Ω denote a discrete set of n vertices (pixels) endowed with a neighborhood relationship $\mathcal{N} : \Omega \to \mathcal{P}(\Omega)$, the power set of Ω. For each pixel $x \in \Omega$, the set $\mathcal{N}(x)$ is the set of all neighbors of x. We assume that the number of elements of $\mathcal{N}(x)$ denoted $|\mathcal{N}(x)|$ is bounded independently of x by a constant $c_{\max}$. In our implementation, we simply use 4-connectivity and therefore $c_{\max} = 4$. We can now define local contrast changes.

Definition 1. *Let u and u_1 denote two images. They are said to differ by a local contrast change if the order relationship between adjacent pixels is the same for each image. More precisely, $\forall x \in \Omega$ and $\forall y \in \mathcal{N}(x)$:*

$$\text{sign}(u(x) - u(y)) = \text{sign}(u_1(x) - u_1(y)), \quad (4)$$

with the convention $\text{sign}(0) = 0$.

[1]We deliberately simplify the definition of level lines here.



(a) Image u_1

(b) Connected components

(c) Associated graph

$$A = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

(d) Transpose of the incidence matrix

Fig. 2: Illustration of the graph construction

B. A compact representation of local contrast changes

Problem (1) can be rephrased as follows:

$$\min_{u\in\mathcal{U}_1} \frac{1}{2}\|u_0 - u\|_2^2, \quad (5)$$

where $\mathcal{U}_1$ is the set of images satisfying condition (4). Let $(\omega_i)_{1\leq i\leq p}$ denote the connected components of the level lines of u_1. These $p \leq n$ components form a partition of Ω. The first step of our algorithm is to construct the sets $(\omega_i)_{1\leq i\leq p}$ and a directed acyclic graph $G = (V, E)$ from the image u_1. The set $V = (v_1, \ldots, v_p)$ are the vertices of this graph and represent the sets $(\omega_i)_{1\leq i\leq p}$. The set $E = (e_1, \ldots, e_m)$ are the edges of the graph. Edge $e_k \in E$ is an ordered pair of vertices written $e_k = (\text{I}(k), \text{J}(k))$ going from vertex $\text{I}(k)$ to vertex $\text{J}(k)$. Such an edge exists if the sets $\omega_{\text{I}(k)}$ and $\omega_{\text{J}(k)}$ are connected via the neighborhood relationship $\mathcal{N}$ and if $u_1(\omega_{\text{J}(k)}) > u_1(\omega_{\text{I}(k)})$. The graph G can be encoded through an incidence matrix (or more precisely its transpose) $A \in \mathbb{R}^{m\times p}$. Each row of this matrix describes an edge with the convention $A(k, \text{I}(k)) = -1$ and $A(k, \text{J}(k)) = 1$ and all the other coefficients of row k are null. A simple 3×3 image u_1, the associated regions $(\omega_i)_{1\leq i\leq 4}$, graph and incidence matrix are represented in Figure 2.

The list of regions $(\omega_i)_{1\leq i\leq p}$ can be constructed in $O(n)$ operations using flood fill algorithms. The graph or matrix A can be constructed in $O(n\log(n))$ operations. The idea is to first scan all edges in $\mathcal{N}$ to construct a preliminary matrix $\tilde{A}$ with repetitions. For instance, region ω_1 is connected three times to ω_3 (see arrows in Figure 2a). The complexity of this part is $O(c_{\max}n)$. Then, the repetitions can be suppressed in $O(n\log(n))$ operations using a quicksort algorithm.

The set $\mathcal{U}_1$ can now be described compactly as follows:

$$\mathcal{U}_1 = \{u : \Omega \to \mathbb{R}, u|_{\omega_i} = \alpha_i,\ 1 \leq i \leq p, A\alpha > 0\}.$$

This set is convex but not closed. The solution of (5) might therefore not exist. To ensure existence, we replace (5) by

$$\min_{u\in\bar{\mathcal{U}}_1} \frac{1}{2}\|u_0-u\|_2^2, \quad \text{where} \tag{6}$$
$$\bar{\mathcal{U}}_1 = \{u:\Omega\to\mathbb{R}, u|_{\omega_i}=\alpha_i,\ 1\le i\le p, A\alpha\ge 0\}.$$

C. A new convex minimization algorithm

In this section, we concentrate on the numerical resolution of problem (6). We first simplify it as follows

$$\min_{u\in\bar{\mathcal{U}}_1} \frac{1}{2}\sum_{x\in\Omega}(u_0(x)-u(x))^2$$
$$= \min_{u\in\bar{\mathcal{U}}_1} \frac{1}{2}\sum_{i=1}^{p}\sum_{x\in\omega_i}(u_0(x)-\alpha_i)^2.$$

Next, let $\beta_i = \frac{1}{|\omega_i|}\sum_{x\in\omega_i} u_0(x)$ denote the mean of u_0 over region ω_i. We get:

$$\sum_{x\in\omega_i}(u_0(x)-\alpha_i)^2 = \sum_{x\in\omega_i}(u_0(x)-\beta_i+\beta_i-\alpha_i)^2$$
$$= \sum_{x\in\omega_i}(u_0(x)-\beta_i)^2+(\beta_i-\alpha_i)^2+2(u_0(x)-\beta_i)(\beta_i-\alpha_i)$$
$$= |\omega_i|\mathrm{Var}_{\omega_i}(u_0)+|\omega_i|(\beta_i-\alpha_i)^2,$$

where $\mathrm{Var}_{\omega_i}(u_0)$ is the variance u_0 over region ω_i. Therefore, the problem reads

$$\min_{A\alpha\ge 0}\frac{1}{2}\sum_{i=1}^{p}|\omega_i|(\beta_i-\alpha_i)^2+|\Omega|\mathrm{Var}_\Omega(u_0).$$

By letting $w\in\mathbb{R}^p$ denote the vector with components $w_i=|\omega_i|$, $W=\mathrm{diag}(w)$ and by skipping the constant term $|\Omega|\mathrm{Var}_\Omega(u_0)$, problem (1) finally simplifies to:

$$\min_{A\alpha\ge 0}\frac{1}{2}\langle W(\alpha-\beta),\alpha-\beta\rangle. \tag{7}$$

Problem (7) is - once again - an isotonic regression problem. It is however much more complicated than problem (3), due to the near arbitrary structure of matrix A. In the rest of the paper, we let α^* denote the unique minimizer of (7). Uniqueness is due to the fact that (7) amounts to projecting a vector onto a *closed convex set*.

Isotonic regression problems received a considerable attention in the optimization literature (see e.g. [11], [12], [16], [17] for a few algorithms). To the best of our knowledge, the best existing complexity estimates to solve them - in the case of an arbitrary directed acyclic graph - were provided recently in [18]. Therein, the authors propose an interior point algorithm [19] exploiting the special graph structure of matrix A [20]. Let α^* denote the unique solution of problem (7). Their tailored algorithm provides a feasible estimate $\alpha^{(\epsilon)}$ of α^* satisfying $A\alpha^{(\epsilon)}\ge 0$ with $\|\alpha^{(\epsilon)}-\alpha^*\|_2^2\le\epsilon$ in no more than

$$O(m^{1.5}\log^2 p\log(p/\epsilon)) \tag{8}$$

operations. In practice, this algorithm works nicely for small m, but in our experience, the method fails to converge when dealing with large graphs. In what follows, we therefore design a simpler first order algorithm.

The main idea is to exploit strong convexity of the squared l^2-norm to design a first order algorithm on the dual. Nesterov type accelerations [21] make this method particularly relevant for large scale problems [22]. Proposition 1 summarizes the nice properties of the dual problem.

Proposition 1. *The dual problem of* (7) *reads:*

$$\sup_{\lambda\le 0} D(\lambda) = -\frac{1}{2}\|W^{-1/2}A^T\lambda\|_2^2+\langle\lambda,A\beta\rangle. \tag{9}$$

Let $\alpha(\lambda)=\beta-W^{-1}A^T\lambda$*, then any primal-dual solution* (α^*,λ^*) *satisfies* $\alpha^*=\alpha(\lambda^*)$*. Function* D *is differentiable with L-Lipschitz continuous gradient and* $L=\lambda_{\max}(AW^{-1}A^T)$*. Finally, the following inequality holds for any* $\lambda\in\mathbb{R}^m$*:*

$$\|\alpha(\lambda)-\alpha^*\|_2^2\le 2(D(\lambda^*)-D(\lambda)). \tag{10}$$

Proof. We only sketch the proof. The idea is to use Fenchel-Rockafellar duality for convex optimization:

$$\min_{A\alpha\ge 0}\frac{1}{2}\langle W(\alpha-\beta),\alpha-\beta\rangle$$
$$= \min_{\alpha\in\mathbb{R}^m}\sup_{\lambda\le 0}\frac{1}{2}\langle W(\alpha-\beta),\alpha-\beta\rangle+\langle A\alpha,\lambda\rangle$$
$$= \sup_{\lambda\le 0}\min_{\alpha\in\mathbb{R}^m}\frac{1}{2}\langle W(\alpha-\beta),\alpha-\beta\rangle+\langle A\alpha,\lambda\rangle.$$

The primal-dual relationship $\alpha(\lambda)$ is obtained by finding the minimizer of the inner-problem in the last equation. The dual problem is found by replacing α by $\alpha(\lambda)$ in the inner-problem.

Function D is obviously differentiable with $\nabla D(\lambda)=-AW^{-1}A^T\lambda+A\beta$. Therefore, $\forall(\lambda_1,\lambda_2)$, we get:

$$\|\nabla D(\lambda_1)-\nabla D(\lambda_2)\|_2 = \|AW^{-1}A^T(\lambda_1-\lambda_2)\|_2$$
$$\le \lambda_{\max}(AW^{-1}A^T)\|\lambda_1-\lambda_2\|_2.$$

Inequality (10) is a direct consequence of a little known result about the Fenchel-Rockafellar dual of problems involving a strongly convex function. We refer the reader to lemma D.1 in [23] for more details. □

Proposition 2. *The Lipschitz constant* L *satisfies* $L\le 4c_{\max}$.

Proof. Notice that $\lambda_{\max}(AW^{-1}A^T)=\sigma^2_{\max}(AW^{-1/2})$, where $\sigma_{\max}$ stands for the largest singular value. Moreover

$$\|AW^{-1/2}\alpha\|_2^2 = \sum_{k=1}^{m}\left(\frac{\alpha_{\mathrm{I}(k)}}{\sqrt{w_{\mathrm{I}(k)}}}-\frac{\alpha_{\mathrm{J}(k)}}{\sqrt{w_{\mathrm{J}(k)}}}\right)^2$$
$$\le \sum_{k=1}^{m}2\left(\frac{\alpha^2_{\mathrm{I}(k)}}{w_{\mathrm{I}(k)}}+\frac{\alpha^2_{\mathrm{J}(k)}}{w_{\mathrm{J}(k)}}\right)$$
$$= 4\sum_{k=1}^{m}\frac{\alpha^2_{\mathrm{I}(k)}}{w_{\mathrm{I}(k)}}$$
$$= 4\sum_{i=1}^{p}n_i\frac{\alpha_i^2}{w_i},$$

where n_i denotes the number of edges starting from vertex i (the outdegree). To conclude, notice that each pixel in region

ω_j has at most $c_{\max}$ neighbors. Therefore $n_i \leq w_i c_{\max}$ and we finally get:

$$\|AW^{-1/2}\alpha\|_2^2 \leq 4c_{\max}\sum_{i=1}^{p}\alpha_i^2 \qquad = 4c_{\max}\|\alpha\|_2^2. \tag{11}$$

□

Problem (9) has a simple structure, compatible with the use of accelerated projected gradient ascents methods [24] described in Algorithm 1.

Algorithm 1 Accelerated proximal gradient ascent method.

1: **input:** initial guess $\mu^{(1)} \in \mathbb{R}^m$, $\tau = 1/L$ and Nit.
2: **for** $k = 1$ to Nit **do**
3: $\quad \lambda^{(k)} = \min\left(\mu^{(k)} + \tau\nabla D(\mu^{(k)}), 0\right)$.
4: $\quad \mu^{(k+1)} = \lambda^{(k)} + \frac{k-1}{k+2}(\lambda^{(k)} - \lambda^{(k-1)})$.
5: $\quad \alpha^{(k)} = \alpha(\lambda^{(k)})$.
6: **end for**

Proposition 3. *Algorithm 1 provides the following guarantees:*

$$\|\alpha^{(k)} - \alpha^*\|_2^2 \leq \frac{2L\|\lambda^{(0)} - \lambda^*\|_2^2}{k^2}, \tag{12}$$

where λ^* is any solution of the dual problem (9)*. The complexity to obtain an estimate $\alpha^{(k)}$ satisfying $\|\alpha^{(k)} - \alpha^*\|_2 \leq \epsilon$ is bounded above by*

$$O\left(\frac{m}{\epsilon}\|\lambda^{(0)} - \lambda^*\|_2\right) \text{ operations}. \tag{13}$$

Proof. Standard convergence results [24] state that:

$$D(\lambda^{(k)}) - D(\lambda^*) \leq \frac{L\|\lambda^{(0)} - \lambda^*\|_2^2}{k^2}.$$

Combining this result with inequality (10) directly yields (12).

To obtain bound (13), first remark that each iteration of Algorithm 1 requires two matrix-vector products with A and A^T of complexity $O(m)$. The bound is then a direct consequence of bound (12) and Proposition 2. □

At this point, the convergence analysis is not complete since $\|\lambda^{(0)} - \lambda^*\|_2$ can be arbitrarily large. In order to compare the proposed first order method with the interior point method from [18], we need to upper-bound this quantity.

D. Complexity analysis

In this paragraph, we analyze the theoretical efficiency of Algorithm 1. We consider the special case $W = \mathrm{Id}$ for the ease of exposition. In practice, controlling the *absolute error* $\|\alpha^{(k)} - \alpha^*\|_2$ is probably less relevant than the *relative error* $\frac{\|\alpha^{(k)} - \alpha^*\|_2}{\|\alpha^{(0)} - \alpha^*\|_2}$. This motivates setting $\epsilon = \eta\|\alpha^{(0)} - \alpha^*\|_2$ in equation (13), where $\eta \in [0, 1)$ is a parameter describing the relative precision of the solution.

Setting $\lambda^{(0)} = 0$ and noticing that:

$$\begin{aligned}\|\alpha^{(0)} - \alpha^*\|_2 &= \|\beta - \alpha^*\|_2 \\ &= \|A^T\lambda^*\|_2,\end{aligned}$$

the complexity in terms of η becomes:

$$O\left(\frac{m}{\eta}\frac{\|\lambda^*\|_2}{\|A^T\lambda^*\|_2}\right). \tag{14}$$

In what follows, we show that - unfortunately - the ratio $\frac{\|\lambda^*\|_2}{\|A^T\lambda^*\|_2}$ can behave like m and the complexity therefore becomes $O\left(\frac{m^2}{\eta}\right)$, which is significantly worst than interior point methods [18], both in terms of dimension and precision. In all practical examples that we treated, the ratio $\frac{\|\lambda^*\|_2}{\|A^T\lambda^*\|_2}$ however remained bounded by values never exceeding 10, explaining the practical success of the proposed method.

a) Example of a hard problem: An example of hard graph (a simple line graph) is provided in Figure 3. For this graph, Algorithm 1 can be interpreted as a *diffusion process*, which is known to be extremely slow. In particular, Nesterov shows that diffusions are the worst case problems for first order methods in [24, p.59].

Proposition 4. *Consider a simple line graph as depicted in Figure 3, with p even and $W = \mathrm{Id}$. Set*

$$\beta_i = \begin{cases} 1 & \text{if } i \leq p/2, \\ -1 & \text{otherwise}. \end{cases} \tag{15}$$

Then the primal-dual solution (α^*, λ^*) of the isotonic regression problem (7) *is given by $\alpha^* = 0$ and*

$$\lambda_k^* = \begin{cases} -k & \text{if } 1 \leq k \leq p/2, \\ -n + k & \text{if } p/2 + 1 \leq k \leq p. \end{cases} \tag{16}$$

This implies that

$$\frac{\|\lambda^*\|_2}{\|A^T\lambda^*\|_2} \sim m. \tag{17}$$

Proof. For this simple graph, $m = p - 1$. To check that (16) is a solution, it suffices to verify that the three conditions (25), (26) and (27) are satisfied. This is done by direct inspection, using the fact that for this graph:

$$(A^T\lambda)_i = \begin{cases} -\lambda_1 & \text{if } i = 1 \\ -\lambda_i + \lambda_{i-1} & \text{if } 2 \leq i \leq p-1 \\ \lambda_{p-1} & \text{if } i = p. \end{cases} \tag{18}$$

Relationship (17) is due to the fact that the sum of squares $\sum_{k=1}^{m} k^2 = m(m+1)(2m+1)/6 \sim m^3$ so that $\|\lambda^*\|_2^2 \sim m^3$ and $\|A^T\lambda^*\|_2^2 = \|\beta\|_2^2 = m$. □

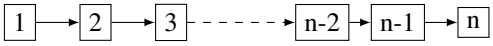


Fig. 3: Worst case graph

b) Example of a nice problem: In order to rehabilitate our approach, let us show that the ratio $\frac{\|\lambda^*\|_2}{\|A^T\lambda^*\|_2}$ can be bounded independently of m for "nice" graphs.

Proposition 5. *For any $\lambda \leq 0$ and the graph depicted in Figure 5 satisfies:*

$$\frac{1}{2} \leq \frac{\|\lambda^*\|_2}{\|A^T\lambda^*\|_2} \leq \frac{1}{\sqrt{2}}. \tag{19}$$

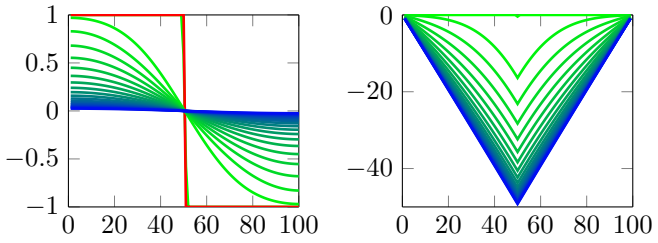


Fig. 4: First 20000 iterations of the primal-dual pair $(\alpha^{(k)}, \lambda^{(k)})$. Top: β is displayed in red while $\alpha^{(k)}$ varies from green to blue with iterations. Bottom: $\lambda^{(k)}$ varies from green to blue with iterations. A new curve is displayed every 1000 iterations. As can be seen, the convergence is very slow.

Proof. For this graph, we get:

$$(A^T\lambda)_i = \begin{pmatrix} -\lambda_1 \\ \lambda_1 + \lambda_2 \\ -\lambda_2 - \lambda_3 \\ \vdots \\ \lambda_{n-2} + \lambda_{n-1} \\ -\lambda_{n-1} \end{pmatrix}. \tag{20}$$

Therefore:

$$\begin{aligned} \|A^T\lambda\|_2^2 &= \lambda_1^2 + \lambda_{n-1}^2 + \sum_{k=1}^{n-2}(\lambda_k + \lambda_{k+1})^2 \\ &= 2\sum_{k=1}^{n-1}\lambda_k^2 + 2\sum_{k=1}^{n-2}\lambda_k\lambda_{k+1}, \end{aligned}$$

and

$$2\|\lambda\|_2^2 \leq \|A^T\lambda\|_2^2 \leq 4\|\lambda\|_2^2. \tag{21}$$

□

In the general case, we conjecture that the method's complexity depends on the length of the longest path in the graph.

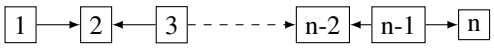


Fig. 5: A nice graph

E. Analytical properties of the minimizer

Let us now provide two analytical properties of the minimizer. The following proposition states that the mean of u^* is equal to that of u_0.

Proposition 6. *The solution α^* of problem (7) satisfies the following property.*

$$\sum_{i=1}^{p} w_i\alpha_i^* = \sum_{i=1}^{p} w_i\beta_i^*. \tag{22}$$

Proof. Let $\mathbb{1}$ denote the vector with all components equal to 1. Notice that $\ker(A) = \text{span}(\mathbb{1})$. Therefore, $\text{Im}(A^T) = \text{span}(\mathbb{1})^\perp$ and we get: $\langle \mathbb{1}, A^T\lambda^* \rangle = \langle \mathbb{1}, W(\alpha^* - \beta) \rangle = 0$. □

The second proposition states that u^* satisfies the *maximum principle* in the sense that:

$$\min_{x\in\Omega} u_0(x) \leq u^*(y) \leq \max_{x\in\Omega} u_0(x) \tag{23}$$

for all $y \in \Omega$.

Proposition 7. *Let $\beta_+ = \max_{1\leq i\leq p}\beta_i$ and $\beta_- = \min_{1\leq i\leq p}\beta_i$, then:*

$$\beta_- \leq \alpha_i^* \leq \beta_+, \quad \forall i \in \{1, \ldots, p\}. \tag{24}$$

Proof. The Karush-Kuhn-Tucker optimality conditions for λ^* read:

$$A^T\lambda^* = W(\beta - \alpha^*), \tag{25}$$

$$\lambda^* \leq 0, \tag{26}$$

$$\lambda_i^* = 0 \qquad \text{if } (A\alpha^*)_i > 0. \tag{27}$$

Now, let $E_{out}(i) = \mathfrak{r}^{-1}(i)$ and $E_{in}(j) = \mathfrak{j}^{-1}(i)$ denote the sets of edges that get in and out of vertex i respectively. With this notation, we get:

$$(A^T\lambda)_i = \sum_{k\in E_{out}(i)} \lambda_k - \sum_{k\in E_{in}(i)} \lambda_k. \tag{28}$$

Let i denote the index of a region corresponding to a local maximum of u_1. This implies that $E_{out}(i) = \emptyset$. Therefore

$$(A^T\lambda^*)_i = -\sum_{k\in E_{in}(i)} \lambda_k^* \geq 0, \tag{29}$$

by equation (26). By equation (25), this implies that $\alpha_i^* \leq \beta_i$.

Therefore, for all positions i corresponding to local maxima of α, we get $\alpha_i^* \leq \beta_i \leq \beta_+$. The fact that all local maxima of α^* are below β_+ implies, in particular, that the global maximum of α^* is below β_+, so $\alpha_i^* \leq \beta_+$ for all $i \in \{1, \ldots, p\}$.

A similar reasoning shows that $\alpha_i^* \geq \beta_-$ for all i. □

F. Analytic properties of SNR_{loc}

To finish this theoretical study, let us mention a few properties of the value $\Delta_{loc}(u_1, u_0)$:

- It is invariant to linear and affine transforms with a coefficient $a \geq 0$, i.e.

$$SNR_{loc}(au_1 + b, u_0) = SNR_{loc}(u_1, u_0), \quad \forall a \geq 0. \tag{30}$$

- It is also invariant to global contrast changes, since global contrast changes are specific instances of local contrast changes. For all non decreasing functions $\phi : \mathbb{R} \to \mathbb{R}$, we get:

$$SNR_{loc}(\phi(u_1), u_0) = SNR_{loc}(u_1, u_0). \tag{31}$$

- In general, it is not symmetric: $\Delta_{loc}(u_1, u_0) \neq \Delta_{loc}(u_0, u_1)$. However, it is possible to make it symmetric by computing $\max(\Delta_{loc}(u_1, u_0), \Delta_{loc}(u_0, u_1))$ or $\min(\Delta_{loc}(u_0, u_1), \Delta_{loc}(u_1, u_0))$.

IV. Robust contrast changes

The proposed methodology has a few properties that may be considered as drawbacks in applications. First, a constant image differs by a local contrast change from *any* other image:

$$SNR_{loc}(constant, u) = +\infty, \ \forall u. \tag{32}$$

This is somehow natural: a photograph taken in a lightless environment could possibly represent any scene. However, it can be a hindrance in some applications.

Another drawback of the proposed approach is illustrated in Fig. 7. In this experiment, we compare two images given in Fig. 7a and 7b of different scenes taken under similar lighting conditions. The minimizer of problem (1) is displayed in Fig. 7c. This image differs from Fig. 7b only by a local contrast change, but is definitely very different. In addition, a very faint gray level variation in Fig. 7b can result in a huge distortion, see red arrows.

All these observations come from the same underlying cause: local contrast changes are a class of transformations that is too wide to correctly model usual lighting variations. In this section, we propose simple restrictions to avoid the mentioned flaws.

A. The principle

The idea is to accept only local contrast variations that belong to a range defined from the contrasts of the input image u_1. This leads to the following definition:

Definition 2. *Let $\theta \leq \Theta$ denote two constants and u and u_1 denote two images. They are said to differ by a robust (θ, Θ) contrast change if the following inequalities are satisfied for all $x \in \Omega$ and for all $y \in \mathcal{N}(x)$ such that $u_1(x) \geq u_1(y)$:*

$$\theta(u_1(x) - u_1(y)) \leq u(x) - u(y) \leq \Theta(u_1(x) - u_1(y)). \tag{33}$$

Notice that θ and Θ are not assumed to be nonnegative, so that this definition also allows for local contrast inversions.

Similarly to the previous sections, let us now consider the following value:

$$\Delta_{rob}(u_1, u_0) = \min_{T \in \mathcal{T}} \|u_0 - T(u_1)\|_2^2, \tag{34}$$

where $\mathcal{T}$ is the set of robust (θ, Θ) contrast changes and define:

$$SNR_{rob}(u_1, u_0) = -10 \log_{10}(\Delta_{rob}(u_1, u_0)/\|u_0\|_2^2). \tag{35}$$

B. A minimization algorithm

Let us express this problem into a form more suitable for numerical computations. Let

$$\gamma_i = \frac{1}{|\omega_i|} \sum_{x \in \omega_i} u_1(x) \tag{36}$$

denote the mean of u_1 over region ω_i. Define $c = \theta A\gamma$ and $C = \Theta A\gamma$. By construction, $A\gamma > 0$ so that $0 \leq c \leq C$, where the inequality is meant component-wise. Instead of looking for the minimizer of (7), we propose to evaluate:

$$\min_{c \leq A\alpha \leq C} \frac{1}{2} \langle W(\alpha - \beta), \alpha - \beta \rangle. \tag{37}$$

In general, the solution of (37) does not satisfy the analytical properties given in Section III-E, but the dual algorithm described in the previous section can be generalized quite easily. The main algorithmic fact are summarized in Proposition 8 and Algorithm 2.

Proposition 8. *The dual problem of* (37) *reads:*

$$\sup_{\lambda \in \mathbb{R}^m} D(\lambda) + \sum_{k=1}^{m} \left(c_k \cdot \min(\lambda_k, 0) + C_k \cdot \max(\lambda_k, 0)\right). \tag{38}$$

Let $\alpha(\lambda) = \beta - W^{-1}A^T\lambda$, then the sequence $\alpha^{(k)}$ generated by Algorithm 2 satisfies:

$$\|\alpha^{(k)} - \alpha^*\|_2^2 \leq \frac{2L\|\lambda^{(0)} - \lambda^*\|_2^2}{k^2}. \tag{39}$$

Algorithm 2 Accelerated proximal gradient ascent method.

input: initial guess $\mu^{(1)} \in \mathbb{R}^m$, $\tau = 1/L$ and Nit.
for $k = 1$ to Nit **do**
  $\tilde{\lambda}^{(k)} = \mu^{(k)} + \tau \nabla D(\mu^{(k)})$.
  $\lambda^{(k)} = \tilde{\lambda}^{(k)} - \tau \min\left(\max\left(\tilde{\lambda}^{(k)}/\tau, c\right), C\right)$.
  $\mu^{(k+1)} = \lambda^{(k)} + \frac{k-1}{k+2}(\lambda^{(k)} - \lambda^{(k-1)})$.
  $\alpha^{(k)} = \alpha(\lambda^{(k)})$.
end for

V. Numerical results

A. Image comparison and change detection

In order to assess the relevance of the proposed approach for image comparison and change detection, we took pictures of two scenes - denoted F and G - under different lighting conditions (window shutter closed or open), see Figures 6 and 7.

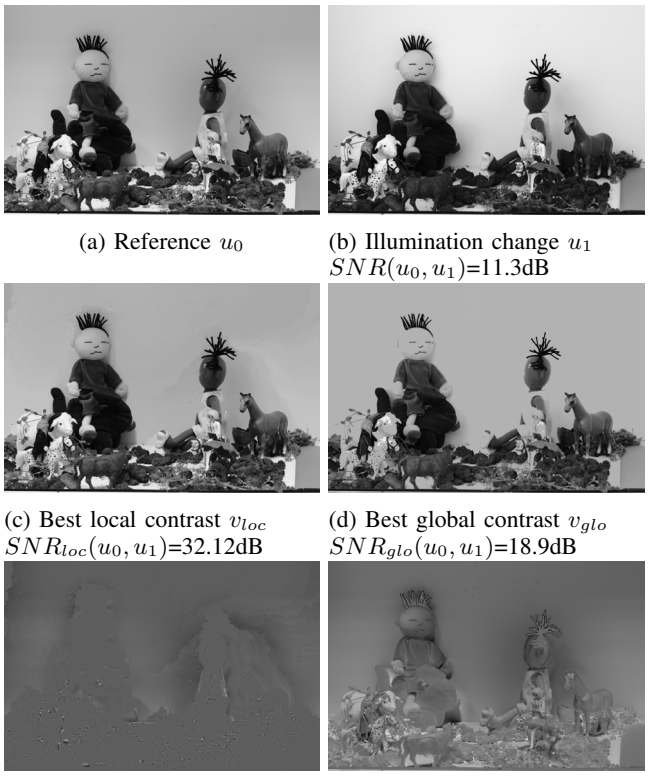
As can be seen in Figures 6 and 7, the proposed local contrast change algorithm is able to correctly distinguish changes of illuminations (Fig. 6e is nearly gray) from changes of scene configuration (Fig. 7e contains only the differences from u to u_1).

B. A large panel of illumination changes

We also evaluated the global and local SNR between all pairs of images in Figure 8. The results are displayed in Tables I and II respectively.

As can be seen in these tables, the SNR between images corresponding to identical scenes is higher than that of images corresponding to different scenes, both for the local and global contrast changes. However this experiment shows that the local SNR is *much more discriminative*. The mean difference between scenes corresponding to different scenes is about 5dB for the global SNR, while it is about 10dBs for the local SNR.

This experiment suggests that local contrast changes provide a much more accurate model of illumination changes than global contrast changes.

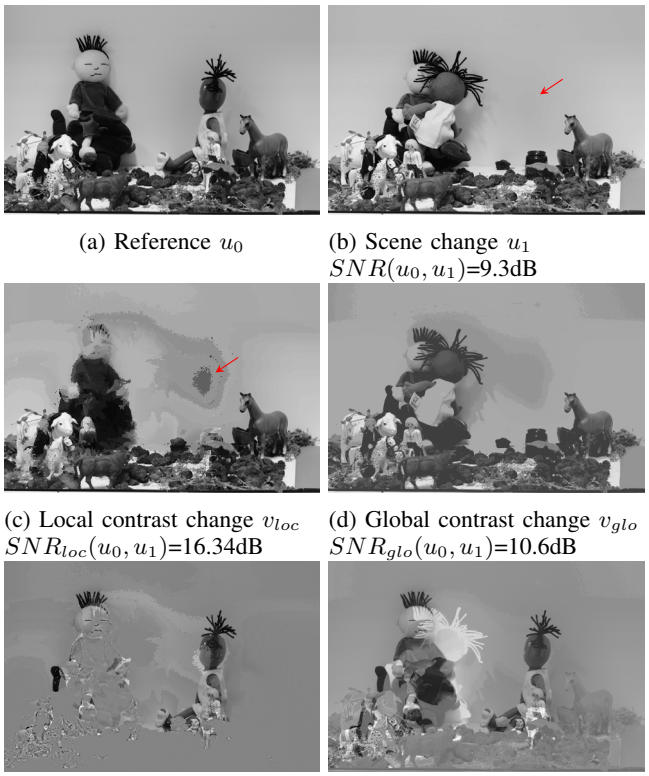


(a) Reference u_0 (b) Illumination change u_1 $SNR(u_0, u_1)$=11.3dB

(c) Best local contrast v_{loc} $SNR_{loc}(u_0, u_1)$=32.12dB (d) Best global contrast v_{glo} $SNR_{glo}(u_0, u_1)$=18.9dB

(e) Difference $u_0 - v_{loc}$ (f) Difference $u_0 - v_{glo}$

Fig. 6: Comparing local and global contrast changes

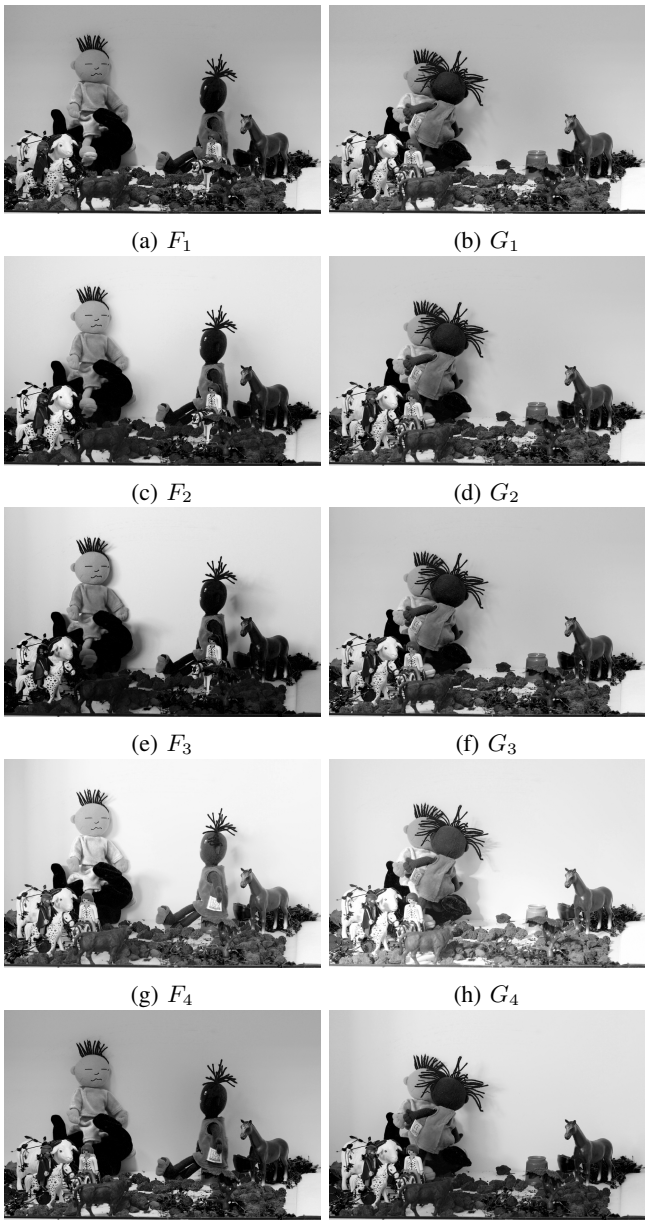


(a) Reference u_0 (b) Scene change u_1 $SNR(u_0, u_1)$=9.3dB

(c) Local contrast change v_{loc} $SNR_{loc}(u_0, u_1)$=16.34dB (d) Global contrast change v_{glo} $SNR_{glo}(u_0, u_1)$=10.6dB

(e) Difference $u_0 - v_{loc}$ (f) Difference $u_0 - v_{glo}$

Fig. 7: Comparing local and global contrast changes



(a) F_1 (b) G_1

(c) F_2 (d) G_2

(e) F_3 (f) G_3

(g) F_4 (h) G_4

(i) F_5 (j) G_5

Fig. 8: Different images used for comparison

	F_1	F_2	F_3	F_4	F_5	G_1	G_2	G_3	G_4	G_5
F_1	Inf	18.11	14.60	16.05	17.24	11.65	11.21	11.14	10.26	12.47
F_2	18.88	Inf	23.91	18.47	15.52	11.51	11.32	11.26	11.00	12.79
F_3	17.05	25.54	Inf	18.45	14.74	11.56	11.40	11.33	11.12	12.87
F_4	14.94	17.38	16.83	Inf	18.04	11.98	11.81	11.75	11.60	13.17
F_5	17.14	15.01	13.03	18.82	Inf	12.24	11.75	11.69	10.64	12.79
G_1	10.54	9.49	8.57	11.44	11.21	Inf	29.56	29.20	18.08	20.20
G_2	10.58	9.68	8.83	11.59	11.24	29.34	Inf	36.34	19.61	21.07
G_3	10.58	9.68	8.82	11.60	11.24	28.81	36.28	Inf	19.76	21.12
G_4	10.41	10.34	9.71	12.46	10.97	19.51	21.59	21.63	Inf	21.49
G_5	10.06	9.48	8.69	11.42	10.58	18.28	19.30	19.30	19.03	Inf

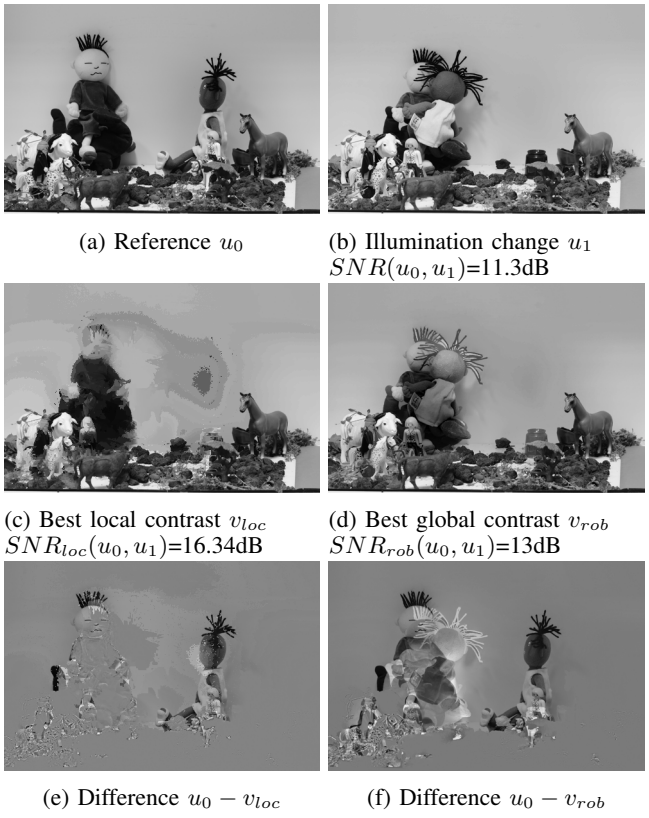
TABLE I: Results global contrasts

	F_1	F_2	F_3	F_4	F_5	G_1	G_2	G_3	G_4	G_5
F_1	Inf	32.12	26.13	25.61	24.64	18.36	18.16	18.10	18.07	19.14
F_2	33.03	Inf	39.26	26.26	23.84	18.36	18.28	18.22	18.34	19.38
F_3	31.28	43.49	Inf	25.86	23.22	17.98	17.80	17.75	17.68	18.93
F_4	23.89	26.00	26.31	Inf	31.17	18.97	18.68	18.65	18.23	19.57
F_5	25.06	25.37	23.42	31.22	Inf	18.89	18.63	18.61	18.52	19.45
G_1	16.91	16.05	14.95	17.98	17.50	Inf	42.15	39.66	31.35	30.53
G_2	16.99	16.36	15.32	18.26	17.65	41.19	Inf	45.04	31.77	31.93
G_3	16.69	16.19	15.20	18.09	17.27	39.07	47.71	Inf	32.56	32.41
G_4	16.10	15.86	15.13	18.01	16.68	31.85	34.76	34.83	Inf	35.04
G_5	15.85	15.45	14.48	17.49	16.33	28.32	31.57	31.67	30.10	Inf

TABLE II: Results local contrasts

C. Robust contrast change

Finally, we show how the robust contrast change described in Section IV behaves in Fig. 9. We set $\theta = 0.5$ and $\Theta = 2$. As can be seen on Fig. 9d, the robust contrast change provides an image that could really be due to lighting variations: it much better resembles the original picture 9b than the local contrast change in Fig. 9c.



(a) Reference u_0 (b) Illumination change u_1 $SNR(u_0, u_1)$=11.3dB

(c) Best local contrast v_{loc} $SNR_{loc}(u_0, u_1)$=16.34dB (d) Best global contrast v_{rob} $SNR_{rob}(u_0, u_1)$=13dB

(e) Difference $u_0 - v_{loc}$ (f) Difference $u_0 - v_{rob}$

Fig. 9: Comparing local and robust contrast changes

ACKNOWLEDGMENT

The authors wish to thank Jonas Kahn for helping finding a hard problem. They thank Pascal Monasse for encouraging them to explore this question. P. Escande is supported by the PRES of Toulouse University and Midi-Pyrénées region. Finally, P. Weiss thanks his daughter Anouk for lending her toys to generate the pictures.

REFERENCES

[1] J. Serra, *Image analysis and mathematical morphology*. Academic press, 1982.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[3] V. Caselles, B. Coll, and J.-M. Morel, "Topographic maps and local contrast changes in natural images," *International Journal of Computer Vision*, vol. 33, no. 1, pp. 5–27, 1999.

[4] V. Caselles and P. Monasse, *Geometric description of images as topographic maps*. Springer, 2009.

[5] P. Weiss, A. Fournier, L. Blanc-Féraud, and G. Aubert, "On the illumination invariance of the level lines under directed light: Application to change detection," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 448–471, 2011.

[6] V. Caselles, B. Coll, and J.-M. Morel, "Geometry and color in natural images," *Journal of Mathematical Imaging and Vision*, vol. 16, no. 2, pp. 89–105, 2002.

[7] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for P+XS image fusion," *International Journal of Computer Vision*, vol. 69, no. 1, pp. 43–58, 2006.

[8] M. J. Ehrhardt and S. R. Arridge, "Vector-valued image processing by parallel level sets," *Image Processing, IEEE Transactions on*, vol. 23, no. 1, pp. 9–18, 2014.

[9] J. Delon, "Midway image equalization," *Journal of Mathematical Imaging and Vision*, vol. 21, no. 2, pp. 119–134, 2004.

[10] A. C. Bovik, *Handbook of image and video processing*. Academic press, 2010.

[11] R. L. Dykstra, T. Robertson *et al.*, "An algorithm for isotonic regression for two or more independent variables," *The Annals of Statistics*, vol. 10, no. 3, pp. 708–716, 1982.

[12] M. J. Best and N. Chakravarti, "Active set algorithms for isotonic regression; a unifying framework," *Mathematical Programming*, vol. 47, no. 1-3, pp. 425–439, 1990.

[13] G. Matheron, *Random sets and integral geometry*. John Wiley & Sons, 1975.

[14] P. Monasse and F. Guichard, "Fast computation of a contrast-invariant image representation," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 860–872, 2000.

[15] L. Najman and M. Couprie, "Building the component tree in quasi-linear time," *Image Processing, IEEE Transactions on*, vol. 15, no. 11, pp. 3531–3539, 2006.

[16] R. Barlow and H. Brunk, "The isotonic regression problem and its dual," *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 140–147, 1972.

[17] P. M. Pardalos and G. Xue, "Algorithms for a class of isotonic regression problems," *Algorithmica*, vol. 23, no. 3, pp. 211–222, 1999.

[18] R. Kyng, A. Rao, and S. Sachdeva, "Fast, Provable Algorithms for Isotonic Regression in all L_p-norms," in *Advances in Neural Information Processing Systems*, 2015, pp. 2701–2709.

[19] Y. Nesterov, A. Nemirovskii, and Y. Ye, *Interior-point polynomial algorithms in convex programming*. SIAM, 1994, vol. 13.

[20] D. A. Spielman and S.-H. Teng, "Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems," in *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*. ACM, 2004, pp. 81–90.

[21] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $o(1/k^2)$," in *Soviet Mathematics Doklady*, vol. 27, no. 2, 1983, pp. 372–376.

[22] P. Weiss, L. Blanc-Féraud, and G. Aubert, "Efficient schemes for total variation minimization under constraints in image processing," *SIAM journal on Scientific Computing*, vol. 31, no. 3, pp. 2047–2080, 2009.

[23] C. Boyer, P. Weiss, and J. Bigot, "An algorithm for variable density sampling with block-constrained acquisition," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 1080–1107, 2014.

[24] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*. Springer Science & Business Media, 2013, vol. 87.